GeoJournal
DOI 10.1007/s10708-007-9103-y

Metrics for the comparative analysis of geospatial datasets with applications to high-resolution grid-based population data

Aarthy Sabesan · Kathleen Abercrombie · Auroop R. Ganguly · Budhendra Bhaduri · Eddie A. Bright · Phillip R. Coleman

Abstract Geospatial data sciences have emerged as critical requirements for high-priority application solutions in diverse areas, including, but not limited to, the mitigation of natural and man-made disasters. Three sets of metrics, adopted or customized from geostatistics, applied meteorology and signal processing, are tested in terms of their ability to evaluate geospatial datasets, specifically two population databases commonly used for disaster preparedness and consequence management. The two high-resolution, grid-based population datasets are the following: The LandScan dataset available from the Geographic Information Science and Technology (GIST) group at the Oak Ridge National Laboratory (ORNL), and the Gridded Population of the World (GPW) dataset available from the Center for International Earth Science Information Network (CIESIN) group at Columbia University. Case studies evaluate population data across the globe, specifically, the metropolitan areas of Washington DC, USA, Los-Angeles, USA, and Houston, USA, and London, UK, as well as the country of Iran. The geospatial metrics confirm that the two population datasets have significant differences, especially in the context of their utility for disaster readiness and mitigation. While this paper primarily focuses on grid based population datasets and disaster management applications, the sets of metrics developed here can be generalized to other geospatial datasets and applications. Future research needs to develop metrics for geospatial and temporal risks and associated uncertainties in the context of disaster management.

Keywords Geospatial data · Population · Statistical evaluation · Disaster management

A. Sabesan
Computational Sciences and Engineering, Oak Ridge National Laboratory, 1 Bethel Valley Road, MS 6085, Oak Ridge, TN 37831, USA

K. Abercrombie · A. R. Ganguly (✉) · B. Bhaduri · E. A. Bright · P. R. Coleman
Computational Sciences and Engineering, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831, USA
e-mail: gangulyar@ornl.gov

Introduction

Accurate estimates of the "population at risk" at high spatial resolutions are important for disaster preparedness, consequence management as well as relief and rescue operations. While grid-based, high-resolution, global population models have widespread use in disaster management, (e.g., Garb et al. 2007; NRC 2007; Rabelo et al. 2006; Dilley et al. 2005), one drawback that limits the utility of such models and corresponding datasets as aids to decision support is the lack of formal approaches for error or uncertainty

quantification. Quantitative assessments of population uncertainties are important to understand and plan for natural or man-made disasters, for example in the context of resource allocations during readiness efforts, as well as intelligent consequence management. This article takes a first step in the direction of formal uncertainty quantification by developing a set of tools to evaluate the utility of high-resolution population data.

Depending on the spatial interpolation algorithm, the national/administrative census population can be represented as continuous gridded population distribution (or count) datasets. The Geographical Information and Science Technology (GIST) group at Oak Ridge National Laboratory (ORNL) reconciles census population counts obtained at relatively low-resolution census polygons, with ancillary information available at high resolutions, to generate high resolution LandScan Global dataset. The LandScan Global 2004 data are available at 30-arc-second spatial resolutions. The ancillary information includes land cover, proximity to roads, and topographical slope obtained from geographical databases and from remote sensing (Bhaduri et al. 2002; Dobson et al. 2000). The Gridded Population of the World with Urban-Rural reallocation (GPW3UR) dataset, for 2000, available from the Center for International Earth Science Information Network (CIESIN) group at Columbia University, relocates the population numbers by distribution within administrative boundaries based on urban–rural extends. The urban–rural extend information is generated by a variety of input data that include census data, online web sources and National Imagery and Mapping Agency (NIMA) database of populated places (Deichmann et al. 2001). The GPW3UR data are also available at 30-arc-second resolutions. Other population datasets from CIESIN include GPW2 and GPW3, which employ areal interpolation to reconcile the population numbers for 1995 and 2000 respectively. The United Nations Environmental Program (UNEP) employs an accessibility potential interpolation to generate grid-based population of the world dataset for 1990 (Deichmann 1996). This article focuses on comparing the LandScan Global 2004 and the GPW3UR 2000 datasets. The primary motivation of this research was to get an answer to the following question: If the most recently updated versions of the two most widely used high-resolution population databases, LandScan Global and GPW3UR, were utilized for disaster preparedness and emergency management efforts, would the differences be quantifiable, and if so, would these differences be significant? The population datasets, both of which represent estimates of the true population, were not compared with ground truth, but with each other. The datasets were of identical spatial resolutions, specifically 30-arc-seconds, hence a spatial alignment was not necessary. Since the purpose was to compare the best available datasets at the time of study, and since comparing with ground truth was not the goal, temporal alignment was not deemed necessary.

Traditional formulations in statistics (Draper and Smith 1998) or geographic information science (Fotheringham et al. 2004; Ripley 2004; O'Sullivan and Unwin 2003; Fotheringham et al. 2000; Chiles and Delfiner 1999; Cressie 1993) can be leveraged for this purpose even though they may not be adequate. Previous studies by Ganguly et al. (2005) have proposed a bottom-up approach for uncertainty quantification, which can combine input- and model-driven (based on individual operation) uncertainties through ensemble simulation approaches and process-based validation. However translating the theory to massive volumes of data and complex processes encountered in the real-world is a significant challenge.

Geo-referenced metrics that can effectively quantify the utility of high-resolution, grid-based population datasets for disaster management do not exist in the literature. In this paper, we investigate the applicability of three sets of metrics, adopted or customized from geo-statistics, applied meteorology and signal processing. These interdisciplinary metrics were then applied to compare the grid-based population datasets. The metrics quantify the difference structures, as well as properties like spatial correlations, and certain features related to exceedence of prescribed population thresholds.

The result of our study indicates that relatively smaller and sharper population clusters characterize the LandScan datasets while GPW3UR datasets exhibit relatively diffuse spatial population clusters. In addition, our results demonstrate that in a majority of cases LandScan and GPW3UR datasets rarely agree on the geospatial placement of high population values. These results and significance are elaborated in the following sections.

Data and metrics

Data

The LandScan Global dataset for the year 2004 was obtained from the ORNL. The GPW3UR (GRUMP alpha dataset) for the year 2000 was downloaded from CIESIN group website at Columbia University (http://sedac.ciesin.org/gpw/global.jsp—Verified July 21, 2007). The GPW3UR data downloaded as ArcInfo Interchange file (.eoo format) was converted to a project standard ESRI grid file. ESRI data on world countries was also used in the study region analysis. A straightforward computation of the geospatial metrics described in this paper requires datasets with identical spatial resolutions. If two datasets with different spatial resolutions need to be compared, they must be brought to identical resolutions through spatial aggregation or disaggregation prior to the analysis. Since both the LandScan and GPS3UR datasets have a 30-arc-second spatial resolution, aggregation or disaggregation was not necessary. In general, while some of the metrics described in this paper may be modified for evaluating population datasets with disparate spatial resolutions, caution is strongly advised, especially when accounting for scale effects.

Metrics

Metrics for aggregate and spatially distributed difference

The first set of metrics, collectively called "difference measures", is well suited for detailed difference analysis. This analysis is well suited for detailed analysis of the difference among geospatial datasets at spatially distributed and aggregate scales. These include spatially aggregated measures like normalized mean squared difference (normalized with the product of the standard deviations of the datasets), fractional area coverage (or the total area of the grids with population above pre-defined threshold values) and grid-based difference measures, with or without transformations like the natural logarithm. These metrics are displayed using traditional error analysis tools. Spatial plots were also developed to visualize the goodness of fit in space, and whether there were any obvious relatively large-scale spatial errors.

Metrics for spatial dependence structures

The second set of metrics comprises spatial auto- and cross-correlation measures as functions of spatial lags. The spatial cross-correlation metrics can also be interpreted as "spatially aware" measures of difference. The spatial correlations in each direction are computed by extending the traditional approaches for calculating autocorrelations and correlations used for time series analysis (Box et al. 1994; Mills 1991), in the context of spatial data. For two spatially distributed variables X and Y, which are available in spatial grids (similar to the "lattice" data of Cressie 1993), the spatial dependence structure between the two variables, as a function of spatial "lags" (distances measured as multiples of grid spacing) in each direction is studied by measuring the spatial linear correlation (Pearson product-moment sample correlation) measure, as in Eq. 1.

$$\rho = \sum_{i=1}^{N}\sum_{j=1}^{N}[(X_{(i,j)} - \mu_x) \times (Y_{(i,j)} - \mu_y)] \Big/ \{N^2 \times [\sigma_x \times \sigma_y]\} \quad (1)$$

where ρ represents the spatial correlation value.

$X_{(i,\,j)}$ and $Y_{(i,\,j)}$ represent the two spatially distributed variables at pixel location (i, j).

μ_x and μ_y represent the mean values associated with the spatially distributed variables X and Y, respectively.

σ_x and σ_y represent the standard deviation values associated with the spatially distributed variables X and Y, respectively.

Equation 1 is utilized to provide estimates of autocorrelation and cross-correlation in space at spatial lag zero, or in the "zeroth" direction (i.e., exact superposition). The autocorrelation can be estimated by making the comparison variable the same as the base variable, while the cross-correlation can be obtained by choosing two different variables. The spatial (auto- or cross) correlations can be calculated as a function of spatial lag and direction by selecting all pairs of points that are equidistant and in the same direction from each other. Thus, adjacent pairs of cells result from the eight compass directions (N, S, E, W and NE, NW, SE, SW). The correlation values can be plotted according to the relative position of the pairs in space and can help identify, among other features, clusters or directionality in the

linear spatial dependence. If the number of samples is identical, a visual estimate of the relative correlations at multiple spatial lags does not require the normalization by the square of the number of samples.

Metrics for geospatial comparison based on exceedence thresholds

The third set of metrics is designed to measure the effectiveness of geospatial data in terms of their end-use (Murphy 1993), specifically the use of population data for disaster management. These metrics combine and refine the concepts of "equitable threat scores" and other skill scores used to rank meteorological or climate predictions (Ganguly and Bras 2003).

Consider the two spatially distributed variables, X and Y. X, the base variable, would be used as the base to calculate threshold range, which is typically 75th and 95th percentiles of X. At each threshold within this range, the X and Y variables are examined on a pixel-by-pixel basis, ignoring spatial orientation. Each pixel in X and Y are compared to the threshold and scored accordingly. When the base variable and the comparison variable both exceed the threshold, the occurrence is termed a "Hit", but when both do not exceed the threshold, the occurrence is called a "Correct Negative". If the base variable exceeds the threshold but the comparison variable does not, the occurrence is a "Miss". If the comparison variable exceeds the threshold but the base variable does not, the occurrence is a "False Alarm" (Stanski et al. 1989). These metrics rely on quantities analogous to the probability of detection, false alarm ratio and the probability of false detection and are explained as below:

1. Proportion Correct (PC) = [Hits + Correct negatives]/Total
2. Frequency Bias Index (FBI) = (Hits + False Alarms)/(Hits + Misses)
3. Probability of Detection (POD) = Hits/(Hits + Misses)
4. False Alarm Ratio (FAR) = False alarms/(Hits + False alarms)
5. Probability of False Detection (POFD) = False alarms/(Correct Negatives + False alarms)
6. Threat score = Hits/(Hits + Misses + False alarms)
7. Equitable Threat Score (ETS) = (Hits – Chance)/(Hits + Misses + False alarms – Chance)
8. Chance = (Hits + Misses) (Hits + False alarms)/Total

These metrics are refined such that they provide measures for evaluating multiple geospatial datasets rather than measures for evaluating prediction skills or signal strengths. The receiver operating characteristic (ROC) curve is used to visualize the relationship between the false alarm rate and the hit rate. This metric is used to describe the underlying associations between the exceeding of the threshold and the deficiency to meet the threshold.

A number of other geospatial comparative analysis metrics were developed to determine the underlying relationships the base and the comparison dataset (Y). The geospatial comparative metrics are described in Table 1.

Results and discussions

The geospatial metrics described earlier were used to compare the LandScan and GPW3UR datasets in specific areas of the following case study regions:

1. Los Angeles, California, USA
2. Washington DC, USA
3. Houston, Texas, USA
4. London, UK
5. Iran

The case study regions were selected based on the following considerations: (a) Primary focus on data-rich, urban and densely populated regions in the USA, resulting in the choice of three of the largest US metropolitan areas, but with wide geographical disparity to avoid undue generalizations; (b) Secondary focus on similar data-rich, urban and highly-populated environments in a non-US country, resulting in the choice of a city in Europe (UK); (c) Tertiary focus on data-poor regions of the globe, especially in developing countries over large areas, resulting in the choice of the entire country of Iran. In addition to the above considerations, the choice of the case study regions were also based on anticipated high risks due to man-made hazards (e.g., Washington, Los Angeles or London), natural hazards (e.g., Houston) or terrorism

Table 1 Geospatial comparison metrics

Geospatial comparison metric	
Mutual exceedence	$L > T \cap G > T$
Mutual non-exceedence	$L < T \cap G < T$
Base exceedence comparison non-exceedence	$L > T \cap G < T$
Comparison exceedence base non-exceedence	$L < T \cap G > T$
Mutual exceedence threat score	$(L > T \cap G > T)/(L > T \cup G > T)$
Mutual non-exceedence threat score	$(L < T \cap G < T)/(L < T \cup G < T)$
Base exceedence comparison non-exceedence threat score	$(L > T \cap G < T)/(L > T \cup G < T)$
Comparison exceedence base non-exceedence threat score	$(L < T \cap G > T)/(L < T \cup G > T)$
Bias score	$(G > T)/(L > T)$

L: LandScan; G: GPW; T: Threshold; >: Greater than; <: Less than; ∩: Intersection (set theoretic), equivalent to "AND", ^; ∪: Union (set theoretic), equivalent to "OR", v

and war (Iran). The choice of the specific case study regions should still be viewed as exemplary rather than exhaustive.

The metrics described in this article were tested on all five case study regions. However, this paper presents a summary of the key results through selected figures and discussions. Detailed descriptions and discussions are available in Sabesan et al. (2006).

Difference measures

The absolute difference between LandScan and GPW3UR datasets for these regions are shown in Fig. 1. A visual inspection of the spatial distribution of the difference values indicate significant differences in population numbers between the datasets in regions of high population, especially in downtown Los Angeles, Washington DC and Houston, USA as well as in London, UK. The histograms of the raw dataset indicate differences in the distribution of the population counts in-terms of high and low values, LandScan having higher population differences compared with the GPW3UR datasets. Figure 2a shows the histogram plots for the Washington DC region. Plots of the raw and transformed difference values led to interesting insights. Figure 2b shows that the absolute differences between LandScan and GPW3UR values tend to lie on a straight line with a 45° slope. The slope indicates that for larger population values the differences tend to get magnified, with the LandScan values being mostly higher than the GPW3UR values.

Spatial correlations

The typical spatial auto and cross correlation structures of LandScan and GPW3UR datasets indicate that the LandScan datasets are characterized by sharper population clusters in the dataset than the GPW3UR dataset. The spatial auto-correlations of the LandScan datasets for Houston are seen in Fig. 3a. The figure indicates that the correlation of population is high for smaller lags but decays rather sharply with distance and direction. These plots indicate the presence of small and sharper clusters in the LandScan dataset. The spatial auto-correlations of the GPW3UR dataset for Houston are shown in Fig. 3b. The plot indicates that the correlation is high at smaller lags and also indicates a smoother decline as the lag increases. This implies stronger direction orientation. These plots indicate the presence of more diffuse clusters than the LandScan datasets. The spatial cross-correlations of LandScan and GPW3UR datasets for Houston are shown in Fig. 3c. The figure indicates that the correlation is high at smaller lags and also indicates a smoother decline as the lag increases. The cross-correlation in the East/West direction indicates an increasing trend beyond the 2nd and 6th lag. The cross-correlation values indicate differences in the datasets. Overall, it can be inferred that the spatial cross-correlation structures between LandScan and GPW3UR datasets indicate positive correlations within the datasets at shorter lags.

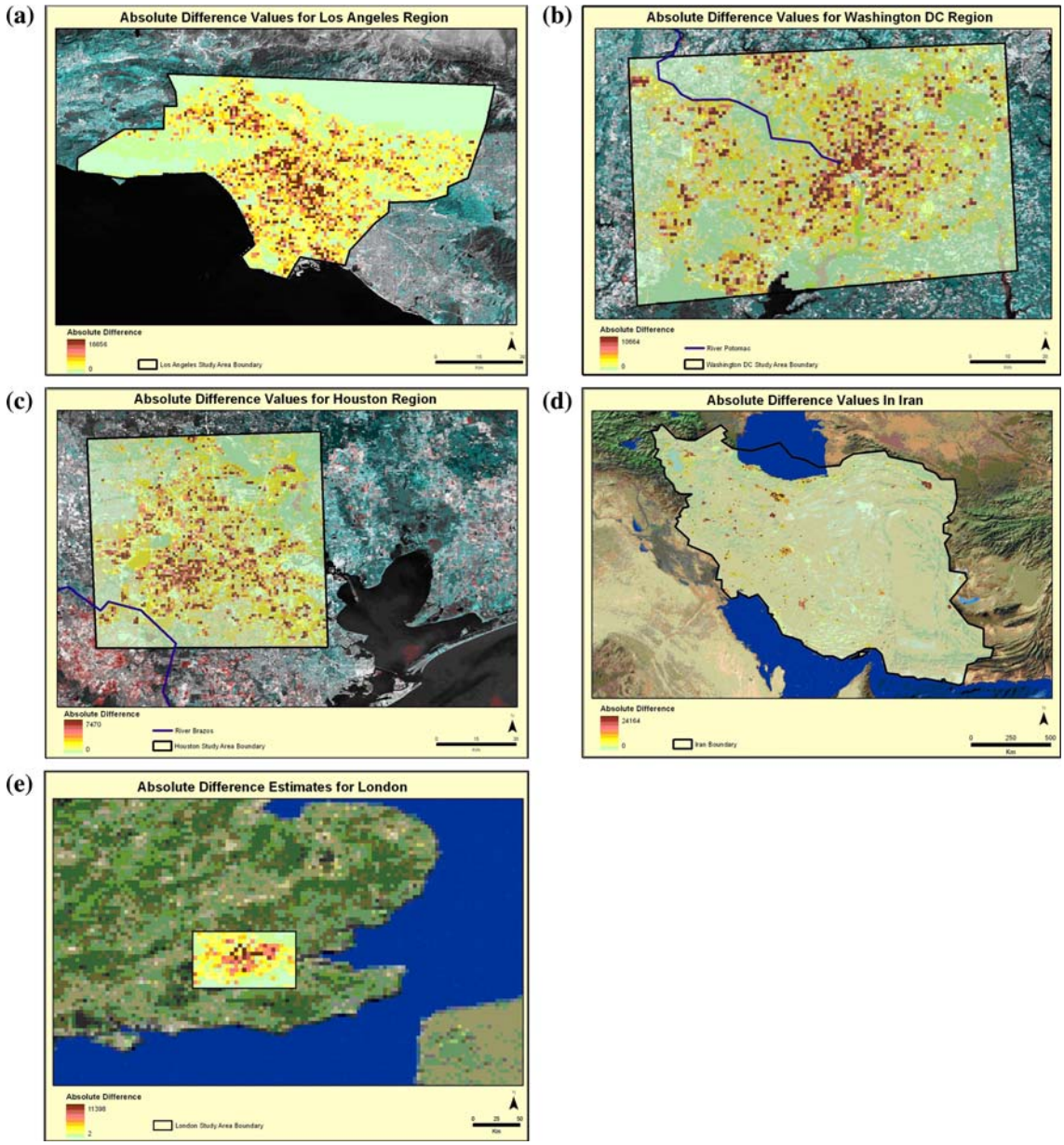

Fig. 1 Absolute difference values between GPW3UR and LandScan datasets for the five study regions

Geospatial comparative metrics

The geospatial comparative metrics were used to show the usefulness of the datasets in case of a natural or man-made disaster, security threat scenarios, or impact analysis. For Houston, the accuracy value (Fig. 4a) decreases to around 70% between approximate threshold values of 200 and 500. The accuracy value drops to a minimum of 65% for thresholds between 600 and 800. Beyond this threshold, the accuracy values increase with increase in threshold, reaching over 90% for population counts beyond 2,000 people. This measure shows the portion of predictions that were the same (either both above the threshold or both below the threshold) increases as the threshold increased. It can thus be inferred that as the threshold increased the ability of GPW3UR and LandScan datasets to predict population

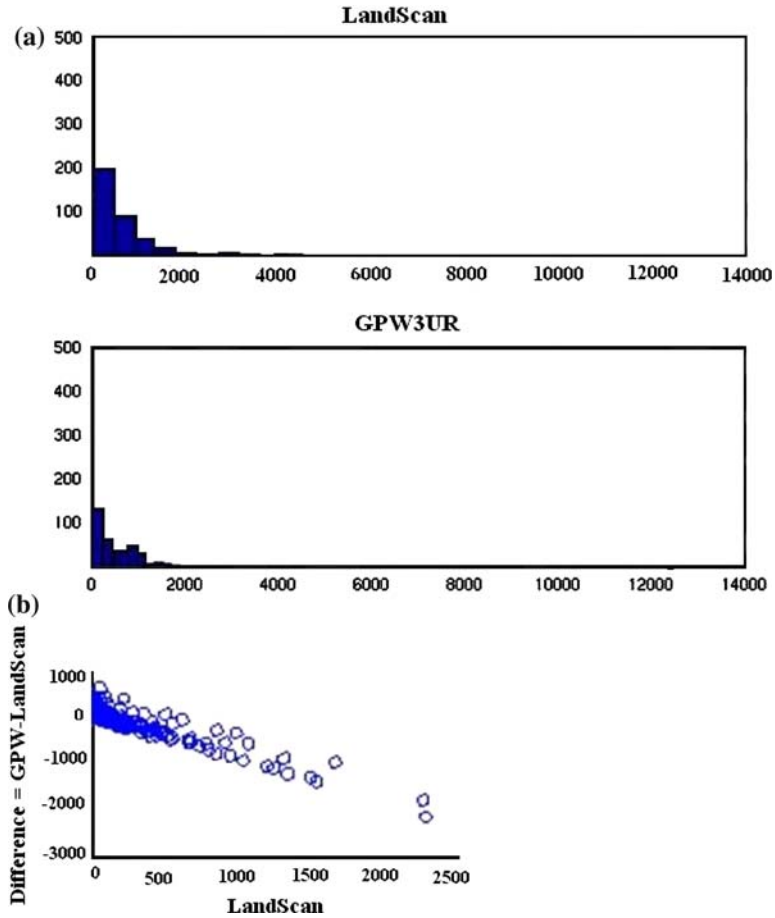

Fig. 2 (**a**) Histogram plots of LandScan and GPW3UR data for Washington DC region (**b**) Different values plotted against the actual LandScan values for the Los Angeles region

increases. For the Los Angeles region, when LandScan dataset is used as the base, the hit rate (Fig. 4b) are relatively high at threshold values around 100 and starts to decline henceforth. Beyond a threshold of 200 people, the POD values tend to be around 0–30%. It is evident that as the population increases, the POD values decreases, indicating increases in the number of "misses". At higher thresholds, the LandScan datasets exceeds the specified threshold, but the GPW3UR dataset does not. At thresholds between 750 and 800, the hit rate is close to zero, indicating that the datasets do not agree on population numbers in this range. The FAR (Fig. 4c) values, for Los Angeles, increase to a maximum of around 0.7 until a threshold of 200 people. The values hover between 0.3 and 0.6 beyond the 200 people. The FAR value for thresholds close to zero approximates 0.3. This indicates that in these population ranges, the GPW3UR exceed the threshold more (around 30%), in comparison to the LandScan datasets.

The threat score values, which ranks the ability of the comparison variable to predict the same populations as the base variable is shown in Fig. 4d for the Washington DC region. The threat score values decrease with the increase in population. The plot indicates that for lower threshold values (0–200), around 70–80% of GPW3UR population estimates were similarly predicted as the LandScan datasets. The threat score values reached a minimum of 20% for threshold values beyond 1,500. It can thus be inferred that at higher populations, the ability of GPW3UR datasets to predict populations similar to LandScan datasets decrease. The mutual exceedence threat score measuring the ability of both datasets to make predictions above the specified threshold is given in Fig. 4e for the Washington DC region. The plot indicates that for lower threshold values, both LandScan and GPW3UR datasets make more predictions above the threshold. This value decreases to reach a minimum of 13% for population ranges

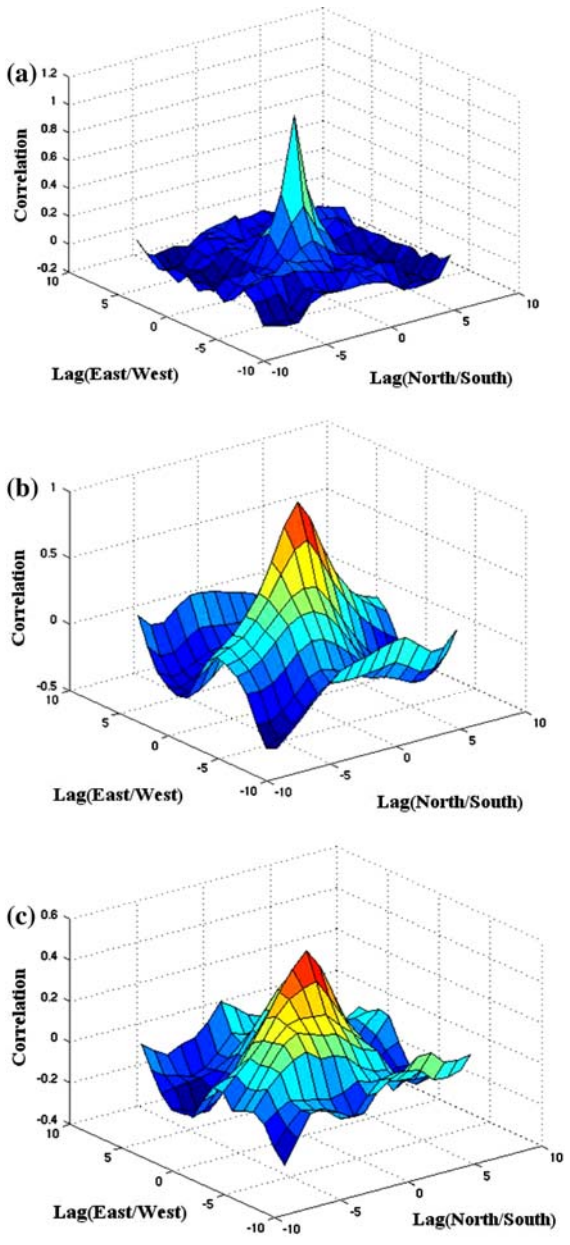

Fig. 3 (**a**) Spatial auto-correlation of LandScan dataset (**b**) Spatial auto-correlation of GPW3UR dataset (**c**) Spatial cross-correlation of LandScan and GPW3UR datasets for the Houston region

between 200 and 250. The mutual non-exceedence plot, for Iran, (Fig. 4f) indicates that as the threshold increases the number of occurrences when both LandScan and GPW3UR datasets do not exceed the specific threshold fluctuate but continues to increase. This confirms our understanding that the LandScan and GPW3UR values do not agree in regions of high population counts.

The Receiver Operating Characteristic (ROC) curve for London (Fig. 4g) shows that the number of hits that occur throughout the range of thresholds is greater than number of false alarms occurring. This implies that whenever the GPW3UR dataset predicts a population exceeding the threshold the LandScan Global dataset will also predict the same exceeding of the threshold.

The results presented here demonstrate the inherent differences among the two population datasets, especially when the population values in one exceed high thresholds. Overall, the LandScan datasets exhibit sharper clusters than the GPW3UR datasets, which in turn have a more diffuse spatial lag structure. The choice of a particular dataset over another may in general depend on the application, even though LandScan datasets do seem to capture greater geospatial variability, which if accurate can be a critical factor for utility in disaster management efforts. In the absence of absolute validation metrics based on ground truths, the facts that the higher variability is caused by additional spatially distributed input variables may be one important consideration for the choice of LandScan over GPW3UR is most disaster management efforts worldwide.

Conclusions and future work

To our knowledge, this is a first attempt to rigorously apply statistical metrics to evaluate high-resolution global-scale population datasets. The two datasets compared, specifically LandScan and GPW3UR, represent the best available population-at-risk estimates currently available to the community. Thus, a direct comparison of the most recent versions of the two datasets is an evaluation of the difference that may result when they are used for disaster preparedness or consequence management. The case studies with the various types of metrics yield several interesting insights regarding the difference of the two population datasets, which have been described earlier.

One important high-level insight is that the LandScan datasets tend to be characterized by relatively smaller and sharper population clusters, while the

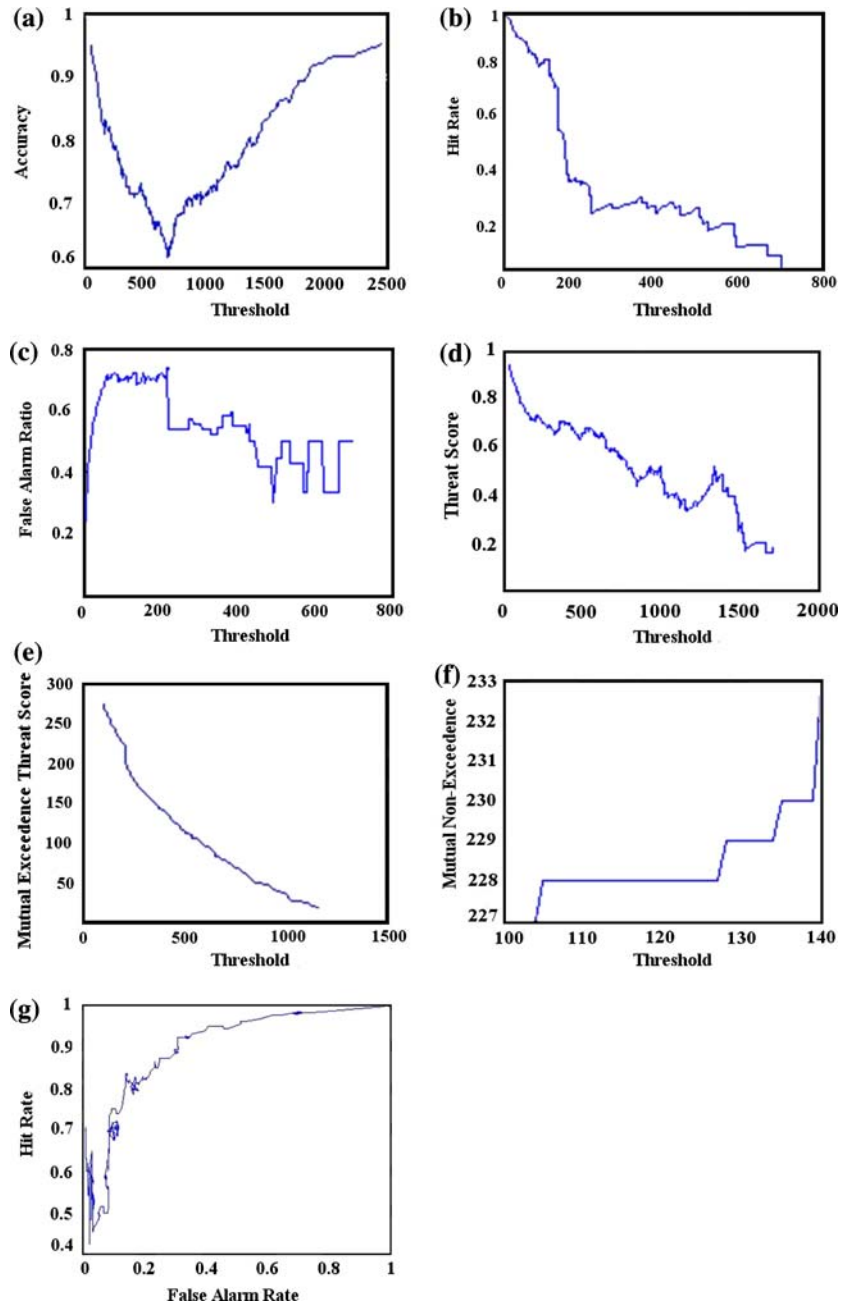

Fig. 4 (**a**) Percent correct score for Houston (**b**) Hit rate score for Los Angeles (**c**) False alarm ratio score for Los Angeles (**d**) Threat score values for Washington DC (**e**) Mutual exceedence threat score for Washington DC (**f**) Mutual non-exceedence plot for Iran (**g**) ROC plot for London

GPW3UR datasets exhibit relatively diffuse spatial population clusters. The fact that LandScan utilizes a more sophisticated and ancillary data-dictated spatial re-distribution algorithm may account for the sharpness of clusters. Generic features such as the tendency to form sharper or more diffuse clusters cannot be taken as a proxy for a measure of accuracy, especially when ground truth is not available for comparison. However, there is evidence from the literature to indicate that populate exhibits clustering tendencies (Small 2004). The ability to faithfully represent observed spatial clusters is important for decision-makers, as the properties of population clusters may dictate the geographic allocation of resources, both for enhanced preparedness and for effective consequence management. Future experiments may be designed to

compare the sharpness of clusters obtained from estimated population datasets with the corresponding clusters obtained from imagery, or even ground truth if available.

A second intriguing high-level insight is that our results suggest that in a majority of cases LandScan and GPW3UR datasets rarely agree on the geospatial placement of high population values. Once again, notwithstanding the sophistication or data-richness of LandScan's redistribution algorithm, the lack of ground truth precludes any generic statement about relative accuracy. However, the placement of larger population counts, or the precise location of densely populated regions, is a key determinant of emergency preparedness and response strategies. In resource-constrained environments, additional resources must be positioned in areas with high population to save as many human lives as possible. Thus, future researchers may need to focus on this important aspect and study the correspondence of the dense population regions estimated by the two datasets with imagery or ground truth if available.

Validation of high-resolution population at global scales remains a challenging problem owing to limited availability of ground truth as well as the difficulty to measure the skills in the estimates. Uncertainties creep in the estimates due to the errors in the inputs, resulting in input-dependent uncertainty, and the subjective nature of the estimation or modeling process, causing process-dependent uncertainty. These issues have been described in Ganguly et al. (2005). A first step to quantification of the uncertainties in population datasets, with a particular emphasis on the utility of such datasets for hazard management, is the development of rigorous metrics for comparative evaluation of the datasets. This paper implements three types of metrics, based on spatially aggregate and distributed difference, spatial dependence structures, and exceedence of population thresholds. The procedures and results suggest the tremendous challenges involved in such formulations, especially when one moves from conceptual and theoretical frameworks to real-world applications. Indeed, the challenges are caused by the nature of the problem and the limitations of the state-of-the-art methodologies. However, based on our results, we believe that the skill assessment problems are potentially solvable, or at least significant improvements are possible, based on recent developments in geographical information science and spatial statistical tools. This paper needs to be viewed as a first in a potentially long line of research in the area. We believe this is a critical area for future research, given the importance of the population estimation, and associated uncertainty quantification, problem in the context of hazard readiness and mitigation (NRC 2007).

Acknowledgments Funding support for developing the LandScan Global and LandScan USA databases is provided by several United States government agencies. The authors would like to thank Amy L. King and internal technical and editorial reviewers at Oak Ridge National Laboratory for their assistance. The proceedings have been co-authored by UT Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains, and the publisher by accepting the article for publication, acknowledges that the United States Government retains, a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

References

Bhaduri, B., Bright, E., Coleman, P., & Dobson, J. (2002). LandScan: Locating people is what matters. *Geoinformatics, 5*(2), 34–37.

Bockarjova, M., Steenge, A. E., & van der Veen, A. (2004). On direct estimation of initial damage in the case of a major catastrophe: Derivation of the "basic equation". *Disaster Prevention and Management, 13*(4), 330–336.

Box, G., Jenkins, G. M., & Reinsel, G. (1994). Time series analysis: Forecasting and control (3rd ed.). Prentice Hall.

Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro Internacional de Agricultura Tropical (CIAT), 2004. Gridded Population of the World (GPW), Version 3. Palisades, NY: Columbia University. Available at http://sedac.ciesin.columbia.edu/gpw/global.jsp (Verified June 2007).

Chiles, J.-P., & Delfiner, P. (1999). *Geostatistics: Modeling spatial uncertainty*. Wiley.

Cressie, N. (1993). *Statistics for spatial data*. Wiley.

Deichmann, U. (1996). *A Review of spatial population database design and modelling. National Center for Geographic Information and Analysis (NCGIA)*. Santa Barbara, CA, USA: University of California, Santa Barbara (UCSB).

Deichmann, U., Balk, D., & Yetman, G. (2001). Transforming population data for interdisciplinary usages: From census to grid. Unpublished manuscript available on-line at: http://sedac.ciesin.columbia.edu/plue/gpw/GPWdocumentation.pdf (Verified: June 2007).

Dilley, M., Chen, R. S., Deichmann, U., Lerner-Lam, A. L., Arnold, M., Agwe, J., Buys, P., Kjekstad, O., Lyon, B., &

Yetman, G. (2005). Natural disaster hotspots: A Global risk analysis, The World Bank, 132 pp.

Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., & Worley, B. A. (2000). LandScan: A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing, 66*(7), 849–857.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis*. Wiley.

Fotheringham, A. S., Brundson, C., & Charlton, M. (2004). Some thoughts on inference in the analysis of spatial data. *International Journal of Geographic Information Science, 18*(5), 447–457.

Fotheringham, A. S., Brundson, C., & Charlton, M. (2000). *Quantitative geography, perspectives on spatial data analysis*. Sage.

Ganguly, A. R., & Bras, R. L. (2003). Distributed quantitative precipitation forecasting using information from radar and numerical weather prediction models. *Journal of Hydrometeorology, 4*, 1168–1180.

Ganguly, R. A., Protopopescu, V., & Sorokine, A. (2005). A bottom-up strategy for uncertainty quantification in complex geo-computational models. *Proceedings of the GeoComputation 2005 Conference*. Michigan: Ann Arbor.

Garb, J. L., Cromley, R. G., & Wait, R. B. (2007). Estimating populations at risk for disaster preparedness and response. *Journal of Homeland Security and Emergency Management, 4*(1), Article 3.

Mills, T. C. (1991). *Time series techniques for economists*. Cambridge University Press, 387 pp.

Murphy, A. H. (1993). What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather Forecasting, 8,* 281–293.

NRC [National Research Council] (2007). Tools and methods for estimating populations at risk from natural disasters and complex humanitarian crises. The National Academies Press, 248 pp.

O'Sullivan, D., & Unwin, D. (2003). Geographic information analysis. John Wiley.

Rabelo, L., Sepulveda, J., Compton, J., Moraga, R., & Turner, R. (2006). Disaster prevention and management for the NASA shuttle during lift-off. *Disaster Prevention and Management, 15*(2), 262–274.

Ripley. B. D. (2004). Spatial statistics. John Wiley and Sons.

Sabesan, A., Abercrombie, K. A., & Ganguly, A. R. (2006). Uncertainty estimates in population distribution models, Oak Ridge National Laboratory Technical Report 2006/540. U.S. Department of Energy.

Small, C. (2004). Global population distribution and urban land use in geophysical parameter space. *Earth Interactions, 8*, 1–18.

Stanski, H. R., Wilson, L. J., & Burrows, W. R. (1989). Survey of common verification methods in meteorology. World Weather Watch Tech. Rep. 8, WMO Tech. Doc. 358, World Meteorological Organization, 114 pp.